

Enduring Emotions and the (Im)Possibility of Forgiveness

Rebecca Chan

San José State University

doi: 10.2478/disp-2024-0015

BIBLID: [0873-626X (2024) 75; pp. 296–317]

Abstract

Once fitting, our reactive attitudes are always fitting—or so this paper argues. For instance, grief and resentment ought to endure eternally since the reasons underlying them (like death and wrongdoing) are permanent. The first part of this paper defends this claim against competing views. The second explores implications of the claim for blameworthiness and forgiveness. Eternally enduring emotions, blameworthiness, and resentment may sound undesirable. However, in addition to providing the best theoretical account of the temporality of our reactive attitudes, my account also produces surprisingly welcome pragmatic upshots for our practices surrounding grief, blame, and forgiveness.

Keywords

blame; fittingness; grief; reactive attitudes; resentment

1 Introduction

A loved one dies, and we grieve. This grief *fits* the loss suffered (i*). Time passes and the reason for grief, *i.e.*, the absence of the loved one, endures (ii*). Yet, grief subsides, and (apparently) fittingly so (iii*). Indeed, one who continues to grieve as intensely as they did when the death occurs might be clinically diagnosed. This generates a puzzle about the

temporality of grief—and the rest of our reactive attitudes: “how could the diminution of grief not be wrong, if [the] reason for grief persists?”¹

There are three general strategies for escaping this puzzle, one corresponding to each of (i*), (ii*), and (iii*), respectively: (i) deny that the reactive attitudes have fittingness conditions, (ii) posit that reasons fade over time, or (iii) hold that our reactive attitudes ought to endure eternally since the reasons underlying them do, too. This paper defends (iii) by arguing against (i) and (ii) (§§2–3), and then offering an error theory that makes (iii) more palatable (§4). It then explores implications for blame (§5) and forgiveness (§6). (Spoiler: Once blameworthy, always blameworthy; and, forgiveness—understood as the releasing of resentment—is impossible.)

Before getting underway, some quick preliminaries are in order. This paper concerns a specific class of our emotions: our *reactive attitudes*. These attitudes are more than mere feeling. They represent the world and are often taken to be evaluative judgments about it. Ideally, grief tells us something good has been lost, resentment signifies wrongdoing, and desire identifies value.² For the purposes of this paper, we can be ecumenical about which emotions count as reactive attitudes and exactly what the attitudes are.³ My core argument focuses on the representational aspect of the reactive attitudes. However, this focus does not preclude the possibility that these attitudes also perform other functions, such as communicative or motivational ones. Indeed, I’ll suggest in later sections that these functions enhance my case for eternally enduring attitudes.

Second, reactive attitudes can be evaluated in terms of *fit*. Fit concerns both *shape* and *size*—i.e., having the correct emotion to the correct degree, respectively.⁴ For instance, take the grief you feel when a loved one dies. That grief is fitting if and only if it has the correct shape and size. Someone who, e.g., intensely grieves when injustices are corrected or when milk spills, does not have fitting attitudes. The fittingness of size is especially important since these features of the world—a loved one dying, corrected injustice, and

¹ Barislav Marušić [2022: 4] lays out and extensively discusses this puzzle. I admire the way he characterizes the puzzle, though I favor a very different resolution.

² See, e.g., Marušić [2022], Strawson [1962] and Oddie [2005], respectively.

³ Other emotions such as anger, rage, gratitude, love, and compassion are often counted as reactive attitudes. Basically, any emotion that’s responsive to moral or value features of the world can count. However, it’s worth noting that some reactive attitudes, such as fear, tend to fit reasons that are temporary. I focus on attitudes corresponding to reasons that are permanent.

⁴ See D’Arms and Jacobson [2000].

spilled milk—are the underlying *reasons* to which attitudes respond.

Finally, this project is essentially *prescriptive*. I claim that our reactive attitudes *ought* to be held eternally, not that they always in fact are. However, this introduces a methodological puzzle. Clearly, descriptive facts about how we in fact hold our reactive attitudes are relevant, and philosophers (including myself) use them to motivate some of our prescriptive claims. (Just think about our original puzzle—(iii*), the claim that it can be fitting for grief to subside, is motivated by the observation that most people experience a subsiding of grief that appears appropriate.) But surely, our prescriptive conclusions should not merely recommend holding our reactive attitudes in the way we already actually hold them.

My methodological approach to these descriptive facts about reactive attitudes is analogous to what I take to be the approach to descriptive facts about belief and action in the context of prescriptive projects about what we ought to believe and what we ought to do. For instance, many people around the world believe that the supernatural exists and take this belief to be rational. Clearly, it would be a mistake to conclude automatically from this fact that we *ought* to believe in the supernatural or that such belief is rational. However, it is appropriate to use this descriptive fact to inform our epistemic theories. For instance, if the vast majority of people believe in the existence of the supernatural, that agreement could offer at least some evidence for the existence of the supernatural and the rationality of holding such a belief. Of course, we'd have to balance that against other descriptive facts, such as disagreement. And, if our epistemic theory ultimately recommends that we ought not to believe in the existence of the supernatural, it ideally should explain why such a widespread belief ought to be rejected as irrational.

Similarly, when it comes to our reactive attitudes, I don't want to say that descriptive facts about the way we hold these attitudes—such as the fact that our grief often fades over time—entail corresponding prescriptive norms—in this case, either that it is fitting for our grief to fade or that our grief ought to fade. (After all, I do reject (iii*), the claim that our grief fittingly fades!) However, these descriptive facts aren't completely divorced from prescriptive norms. They might shed light on what the attitudinal norms are, and it could be appropriate to take these descriptive facts to favor a normative theory that prescribes holding attitudes in that way. Of course, these descriptive facts must be balanced against other descriptive facts that might support that reactive attitudes ought to endure eternally. And, if (and when) it turns out that our theory of reactive attitudes recommends that we ought to hold certain attitudes eternally, that theory ideally should explain why the fading of attitudes like grief is apparently fitting.

2 Fittingness

Without fittingness conditions, reactive attitudes are free to fade.⁵ They might also strengthen or never occur at all. The puzzle dissolves since our attitudes are disconnected from the reasons underlying them.

It's tempting to take this route if one holds an impoverished view of emotions as merely something that we *feel*. Often, they lie outside of our rational control, and it doesn't seem quite right to criticize people's feelings. Nevertheless, emotions bear an interesting analogy to belief. Like our emotions, our beliefs often lie outside of our control. Yet, beliefs still can be "fitting", such as when they match the evidence available to us, and "unfitting", such as when they don't match available evidence. Similarly, the fact that an attitude is not under our full control is compatible with the fact that it's still subject to fittingness norms.

Furthermore, reducing emotions to mere feeling unnecessarily strips them of important moral, social, and psychological roles. For instance, Strawson [1962] famously claims that resentment identifies agency and blameworthiness. In a similar vein, Oddie [2005] argues that we perceive value in the world through our desires. The fruitfulness of reactive attitudes spill into the political domain as well. Srinivasan [2018] and Cherry [2020] both defend anger and rage as fitting responses to injustice that can both highlight an injustice and motivate a response. McSweeney [2022] and Arpaly [2005] point out that even negative directed emotional states such as depression and trauma responses can tell us important things about the world; understanding that they can be 'warranted' (i.e., fitting) carries important implications for any kind of therapy seeking to alleviate these emotions without appreciating the fact that they are warranted responses to the world.⁶

This richer conception of emotions reveals that emotions work in tandem with beliefs to provide us with important information about the world. Ideally, reactive attitudes are *informative* as well as evaluative. Think about somebody you love. You propositionally know *that* you love them. But you presumably also *feel* love for them. This feeling informs your

⁵ Strictly speaking, if reactive attitudes just are emotions that have fittingness conditions, then this amounts to saying that there are no reactive attitudes, just emotions that aren't evaluable for fittingness.

⁶ McSweeney's concept of warrant is, roughly, the equivalent of the fittingness of emotions for directed emotional states. For her "[w]arranted directed emotional states are states that correctly match emotions to their objects" ([2023: 846]). Arpaly's concept of warrant is closer to truth-tracking. Importantly for both, directed emotional states correspond to the world in an informative way, and thus mental illnesses composed of these states cannot be treated in the same way as physical illness.

understanding of them and your love at least as much as any propositional knowledge—it's a sort of direct acquaintance. Never feeling love is defective even if one cognitively knows that their beloved is important in the relevant way.

Indeed, this point features prominently in discussions of other attitudes. Take, for instance, what Srinivasan [2018] says about anger:

I want to suggest that getting angry is a means of affectively registering or *appreciating* the injustice of the world, and that our capacity to get aptly angry is best compared with our capacity for aesthetic appreciation. Just as appreciating the beautiful or the sublime has a value distinct from the value of knowing that something is beautiful or sublime, there might well be a value to appreciating the injustice of the world through one's apt anger—a value that is distinct from that of simply *knowing* that the world is unjust... [2018: 132]

For Srinivasan, anger, like other reactive attitudes, responds to a feature of the world and tells us something that goes beyond propositional knowledge. In fact, lacking the capacity for anger and the like reveals a defect. Mengzi offers a similar example involving a child falling into a well. Lacking concern—even if one cognitively processes that this is a bad situation for the child and takes the appropriate action—is inferior to both having concern and knowing that concern is warranted. People experiencing cognitive dissonance suffer from this defect. They cognitively recognize, *e.g.*, that injustice warrants anger, but either lack anger or feel an unfitting emotion instead. (And, note that common phenomena like cognitive dissonance only make sense given that emotions ought to fit!)

Reactive attitudes are akin to non-propositional knowledge and contribute to our epistemic success. Like beliefs, they fit when they correspond to the world, including the reasons underlying our emotional responses. This focus on the epistemic-like aspect of emotions is not to deny that emotions may have other functions. Plausibly, anger communicates to others that injustice has occurred and also motivates those who feel anger to correct that injustice. And while this paper focuses primarily on the epistemic aspect of emotions, it also has something to say later on about the communicative and motivational aspects as well.

3 Permanence

Perhaps reasons fade over time. This could happen in several different ways: the reason itself could literally disappear, diminish, or change. I'll argue that even if we grant the

metaphysical claim about the reason itself changing, the change in reason either fails to explain the reasonableness of other common attitudinal experiences or fails to explain why attitudes ought to change (§3.1). This leads to a more general point: the temporal distance between us and the reasons underlying our reactive attitudes does not affect the fittingness of our attitudes (§3.2).

3.1 Impermanent reasons

Let's turn first to the suggestion that the reason itself might be impermanent. There's a very literal way in which this might occur. On presentism, only the present time exists; past and future ones do not. So, reasons exist ephemerally in the present, and then cease to exist. Naively, we might be tempted to expect that emotions evaporate with the reasons because there wouldn't be anything in the world for them to respond to. But this obviously can't be right. Setting aside the possibility that presentism is false, this naïve view doesn't explain many of our emotions that appear fitting. For instance, suppose my beloved cat dies. I still continue to love her, and apparently fittingly so, even though on presentism, the reason for my love, my cat, no longer exists. *Mutatis mutandis* for the love you still hold for your dearly departed. Similarly, your breaking a promise to meet me for lunch ten minutes ago is a reason for me to resent you, albeit slightly. On presentism, that reason no longer exists. Nevertheless, it still appears appropriate for me to resent you. Furthermore, something similar applies to future-oriented attitudes. Anyone who has witnessed a loved one die gradually knows that grief can start before the actual death and continues long after it. Generally speaking, regardless of whether they exist, past and future events affect our present attitudinal states. Grief can be appropriate even if the event, and thus reason, toward which it is oriented, strictly speaking, does not exist.

There's an analogue to belief here, too. Presentists face a truthmaking challenge when it comes to explaining how claims about the past can be true. For instance, "there used to be dinosaurs" seems true, but if the world doesn't contain any dinosaurs, then it seems that nothing in the world makes that claim true. But even in these cases, presentists will have a work-around that preserves truths about the past. For instance, they might say something like "it WAS the case that there are dinosaurs" where WAS is a primitive past-tense operator. The main thing to see is that even if *per improbable*, reasons can go out of existence, the (non-)existence of a reason won't account for whether an emotion is (un)fitting.

Alternatively, some people hold that reasons themselves are time-sensitive. For example, perhaps the shock of the loss of a loved one is what causes initial large-sized grief. That's

large-sized grief is fitting, but as the shock wears off, it's (apparently) reasonable for the grief to fade. Something similar might apply if, *e.g.*, one is inclined to think of grief as a process. As the process unfolds, perhaps through the stages of denial, anger, etc., one's grief evolves until it becomes diminished at the end of the process.

I must confess, I do not see how either of these scenarios demonstrate that reasons change or are time-sensitive. Rather, what these cases suggest is that we relate to reasons in particular contexts. In the shock case, we have one reason for grief—the death itself—plus the surprise of the news. Notice that unlike grief, surprise is by its nature temporal. Thus, if we think that the totality of grief corresponds to the combination of the loss itself plus the surprise, it's unsurprising that as the surprise diminishes, the totality of grief does as well. But this does not mean that the grief that corresponds to the loss has diminished! Perhaps one might complain that surprise itself does not generate grief. Instead, it might be a circumstance or background condition that intensifies the feeling of grief. This picture is compatible with the reason for grief remaining constant. What it actually shows is that surrounding circumstances that also affect our experience of grief, and these circumstances can change. Importantly, these additional effects are not part of the reactive attitude responding to the loss itself, which is evidenced by the fact that people can fittingly grieve to different degrees when exposed to the additional effects. For instance, people who are insensitive to surprise might settle quickly into a fitting level of grief that endures, while those who are sensitive to surprise will experience the sharp uptick in initial grief. The effects of the shock eventually taper off, which makes it appear that grief fades.

Something similar applies to the “grief is a process” view. Na'aman [2021] offers the most comprehensive case for the process view and argues that reasons are “rationally self-consuming.” In a nutshell:

[The] “duration of the emotion is part of the background conditions that explain why the fact that made the emotion fitting in the past no longer makes it fitting, or no longer makes fitting the same intensity of emotion. [Na'aman 2021: 252]

For instance, the fact that you've grieved the death of a loved one for months and gone to a therapist to process that grief makes it the case that feeling that grief to the same intensity, or perhaps at all, is not fitting. Grief, and other emotions, fittingly evolve.

However, there are problems with Na'aman's view. Na'aman maintains that the objects of our attitudes, *i.e.*, the reasons corresponding to them, remain tragic in the case of grief, amusing in the case of amusement, shameful in the case of shame, etc. So, he essentially

does hold that reasons are permanent. Nevertheless, even though the object remains as it is, Na'aman claims it is no longer fitting to feel grief, amusement, shame, and the like. So, it turns out that the process view, as presented by Na'aman, calls fittingness into question. As Howard [2022] points out, this amounts to rejecting our fundamental notion of fit:

He [Na'aman] must reject, for instance, that something is fitting (for me) to regret just in case it's regrettable (for me) and that a person is fitting to blame just in case they're blameworthy. [2022: 91]

Marušić [2022] outlines a further problem with Na'aman's view. While Marušić concedes that Na'aman's process view offers plausible descriptive account of our emotions, it does not *explain* why they ought to subside. As he puts it, his "progression in the process of grief does not make the diminution of [his] grief more intelligible", which is what one would expect if emotions ought to diminish given time ([88]). This criticism intersects with Howard's in revealing that the process view is explanatorily unsatisfactory, at least as far as fit is concerned. While the process view does explain the empirical facts regarding the diminution of grief, it does not explain why such diminution is fitting, and indeed ends up rejecting fittingness.

Perhaps the process view is so compelling that we ought to accept its verdict that grief ought to diminish even in the absence of a satisfactory explanation as to why that is. Indeed, as a descriptive fact, many people report an overall diminution of grief over time. Furthermore, there's something that seems amiss if a person who has gone to therapy for years to process their grief yet still feels the initial intensity of grief all the time. (And I'll have something more to say about this later in §4.)

But there are other explanations of what happens when one processes grief that (a) account for why in fact the total sum of our grief often diminishes even though the underlying reasons are permanent and (b) maintain a commitment to fittingness. Just consider what often happens when one processes grief. When a loved one dies, we need to adjust to a new life without this person. This is an especially difficult task if you were close to them and interacted on a regular basis. For instance, if your days revolved around your loved one, you'll now have to navigate your days alone. Or, if you spent a prolonged period of time caring for an ailing loved one, you might feel relief, or even puzzlement about what to do with newfound hours in the day. As time passes, these reasons may no longer persist: you figure out how and with whom to pass your time. Something—perhaps even just the reminder of loss—changes. This might make it appear that on the whole, grief subsides.

However, at least one impermanent reason remains: that your loved one is dead. And that reason still makes fitting grief *of the same magnitude with respect to that reason*. But since the other reasons may no longer apply, if you had felt grief with respect to them as well, then the total amount of grief you feel fittingly changes. Like the shock case, the process case reveals that death reveals us to a host of reasons that correspond to many different attitudes.

Finally, some philosophers make the interesting suggestion that the past, and thus our reasons, can literally be changed. For example, Lebens [2023] argues that the Judaic sense of atonement is best understood as one on which God literally erases past misdeeds. The past is changed such that the sin never occurs. Similarly, some views of forgiveness (e.g., Arendt [1958]) hold that forgiveness can alter the ethical significance of the past, which might also alter the relevant reasons for resentment. I'm skeptical that it's possible to change the past in this way. However, even if it is, many paradigm cases will not be of this sort. Tragically, the past can't be changed such that a loved one is restored to us. And unfortunately, not all wrongdoers see the error in their ways and commit to repentance and atonement. So, we'll still have many reasons that have not been changed, and if our emotions corresponding to them ought to fade, this account won't explain why since those reasons remain constant.

3.2 Why temporal distance doesn't matter

It's tempting to think that something about our *temporal distance* from underlying reasons matters for how strongly we hold our reactive attitudes. As distance increases, the strength of the reason diminishes, so in order to remain fitting, the strength of our attitudes also diminish. This explains the sentiment behind "time heals all wounds" and why we expect grief to fade. It also explains why the enduring intensity of emotions is unfitting: Ahab's anger might have been appropriate immediately after his encounter with Moby Dick, but the everlasting intensity of his anger is unreasonable and constitutes his central character flaw.

However, there are at least two strong positive reasons to remain temporally neutral with respect to the fittingness and intensity of emotions. First, the temporal distance argument has a spatial analogue: people feel compelled to aid a child drowning in a shallow pond in front of them, but not starving children thousands of miles away. Concern decreases as spatial distance increases. But as Singer points out, we *ought* to feel concern for children

thousands of miles away because distance is not morally significant.⁷ Similarly, we *ought* to feel, e.g., grief for lost loved ones regardless of temporal distance because distance doesn't alter the tragedy of their deaths.⁸

Something similar plays out in Parfit's examples regarding past and future pain.⁹ Most people prefer enduring a more painful surgery in the past to a less painful surgery in the future. However, that preference vanishes in the third person case. Like the spatial example, the third-person case suggests that we have deeply ingrained temporal biases that affect our judgments. But once we recognize these biases, we realize that temporal neutrality is highly plausible.

Second, temporal distance doesn't explain anachronistic attitudes. Even when a death drifts further into the past, grief sometimes overwhelms us as sharply as before. Anniversaries of death provoke sharp grief, as can visiting a significant place, or catching a familiar scent. Furthermore, people sometimes grieve anticipatorily. This might happen if a loved one is terminally ill, but it also occurs when the possibility of a future without one's beloved is raised in a particularly compelling way. Temporal distance fails to explain how these anachronistic attitudes are fitting since the strength of, e.g., grief isn't continually diminishing. Perhaps, instead, the emotions are always there, at least dispositionally. And when they become occurrent, they are fitting regardless of *when* they are felt because they are directed at a reason that still hold the same force it always has.

⁷ Singer [1972]. Marušić [2022] also discusses this analogy, suggesting that requiring maintaining the same strength of attitudes is too demanding.

⁸ An anonymous referee raises a potential disanalogy between the spatial and temporal cases that arises out of our ability to intervene: "if someone is starving 100 miles from here, then there is potentially still something I can do about it, whereas if someone is starving 100 years ago, then there is not". The thought is that our ability to do something about something is relevant to how we ought to feel about it. This point connects nicely with the idea that reactive attitudes have a motivational as well as representational function. If you think that attitudes motivate us to action, and that there is no action to which to be motivated in the past temporal case, then perhaps you might arrive at the conclusion that our attitudes toward present and past starvation ought to differ. But this seems mistaken. When one watches, e.g., *Grave of the Fireflies* and learns about how orphaned children in post-WWII Japan starved to death, they ought to feel outrage and grief. Furthermore, even if there's nothing I can do about the children starving in post-WWII Japan, feeling outrage at their plight can motivate me to do something about similarly situated children in, e.g., present-day Palestine as well as attempt to prevent future starvation.

⁹ See especially Parfit [1984: ch. 8].

4 Enduring emotions

Only (iii), the claim that our reactive attitudes ought to endure eternally at the same strength since the reasons for them do as well, remains. This eternal endurance—e.g., that our grief, if fitting, is always fitting at the same strength—flies in the face of how people *actually* experience emotions and what we commonly judge to be fitting. Many people don't grieve indefinitely; they often move on surprisingly quickly. Indeed, as proponents of process views love to point out, people who continue grieving at the same intensity for prolonged periods are clinically diagnosed with depression. The same holds for other emotions. Ahab's enduring anger is pathological—precisely because it never diminishes. We don't often observe enduring emotions, and when we do, they often seem inappropriate, or at least striking. Thus, opting for (iii) faces the challenge of explaining why “moving on” is so common and widely accepted. In this section, I offer an error theory that both explains this common response and makes accepting (iii) more palatable.

In a nutshell, when an attitude fits, it *always* fits to the same degree because the reason is permanent. (Similarly, when a belief is true, it *always* “fits” because the truth of the proposition is permanent.) In the ideal case, then, attitudes (at their fitting size) last eternally. I've emphasized with bold and italics the plural nature of ‘attitudes’ because it's crucial to remember that we have many fitting attitudes. Grief fits the loss of a loved one; love fits all the endearing qualities that belong to that loved one; gratitude fits the blessing of having found and enjoyed a deep, meaningful relationship with that loved one, and so forth. The totality of your experiences with that loved one provides myriad reasons to feel a whole range of attitudes—eternally. And, of course, there are far more reasons in the world than the ones concerning that beloved. Rage fits the starvation and murder of thousands of children across the world, while resentment fits the blameworthiness of the perpetrators of those atrocities. In short, the world contains many reasons to feel diverse emotions.

But we are psychologically limited beings. We're simply not capable of attending to all of our reasons and attitudes at once, just as we're not capable of occurrently believing

all the propositions we know to be true at once.¹⁰ Our attitudes *appear* to fade because we have other things to attend to. In some sense, receding attitudes are appropriate, since maintaining them would entail ignoring other important aspects of our lives. Given that we have so many reasons and reactive attitudes in our lives, being consumed by just one reason and feeling just one attitude, is a failure—even if the attitude is fitting. But attitudes are still dispositionally there, and when we recall the underlying reason, it will always be fitting to, e.g., grieve as strongly as was initially fitting.

This points to the existence of meta-norms that guide our attention toward the reasons around us. I won't offer a full account of these norms here, but generally speaking, they require us to have some perspective about the myriad of reasons and attitudes available to us. We ought not focus single-mindedly on, e.g., a reason for grief. Similarly, we ought not focus single-mindedly on a reason for love, as many tend to do when in the initial stages of infatuation. In these cases, the attitude itself is appropriate because it fits the corresponding reason. The mistake lies in narrowly focusing on that reason to the point that it blinds us to other reasons, thus crowding out other appropriate attitudes.

This error theory also explains some of the attitudinal phenomena discussed in the previous section. In general, we tend to experience our attitudes most strongly when we become acquainted with their underlying reasons. Because we experience time linearly, that is when they typically dominate our attention and explains why we tend to have strong temporal biases. Recognizing this provides an alternative explanation to the most compelling aspects of the process view. When one takes time to process grief, perhaps by going to therapy, they are not changing the fittingness of their original intensity of grief. Instead, they are shifting their perspective and attention so that they can appreciate other reasons associated with their loved one and the rest of their life more generally. In addition to the loss, reasons associated with love and gratitude may resurface. While the grief is

¹⁰ Here's an example illustrating the analogy to belief. My brother hates clams because when we were kids, he once got sick after eating them. I believe this fact. But this belief isn't particularly relevant to my life. I don't often eat with my brother because we live in different states, and there are many more important facts about my brother's life (like the goings on of his children and his obsession with pickleball) that occupy my occurrent beliefs about him. In fact, even when I do see my brother, I sometimes forget about this hatred of clams when we do eat together and I'll propose ordering a dish with clams. It might very well appear that my belief in this fact has receded. But as soon as I'm reminded of his hatred of clams, my belief is as strong as it was before. In a meta sense, given the limits of my mental capacities, it's good that I don't fixate on this one belief and instead focus on others. I hold this belief dispositionally and when it's raised in the right way, I still believe it as strongly as I did before.

there, it comes along with these positive attitudes associated with the loved one that will endure forever as well. There is something amiss if one experiences their original intensity of grief forever, but the intensity of grief is not the problem. Rather, it's the inability to appreciate other reasons and feel other attitudes.

Anachronistic attitudes offer further support for this point. Careful examination of grief experiences reveals that grief sometimes overwhelms us as sharply as before when our attention is turned to the underlying events by, e.g., anniversaries or familiar settings. The same is true of moments that remind us of reasons for love and gratitude toward our loved ones. When our attention is turned toward these reasons, the corresponding attitudes become occurrent rather than dispositional. In these moments, we often do feel grief, love, and gratitude at their original intensity, and that's fitting. (Notice that it wouldn't be if reasons change over time.) This fits the error theory. We do still hold our attitudes as intensely as before, but the underlying reason must be brought back into our attention.

There's a final possibility that this last example and our error theory raises. Thus far, we've assumed that the initial, 'large'-sized attitudes are fitting, and that we needed to account for their diminishment. Given what's been said, it's possible that the initial way we feel these emotions are disproportionately large. For instance, we speak of people being blinded by grief, anger, love, and desire. These attitudes have a tendency to crowd out all others when the reasons they fit first come to light, and we become fixated on them. In these situations, attitudes are outsized, for they blind us to the other reasons we have to feel other emotions. Over time, they recede enough to be felt alongside other emotions. For instance, an initial, overwhelming grief might eventually settle alongside enduring love and gratitude. This picture is compatible with one on which attitudes are fitting and ought not fade. They also ought not be felt too strongly in the first place.

5 Blame

The enduring fittingness of reactive attitudes carries implications for blame. Once it's fitting to resent someone, it is always fitting to resent them! And, *à la* Strawson, since fitting resentment indicates blame, blameworthiness lasts forever as well! Indeed, the more popular view appears to be that resentment, blame, and associated judgments like guilt are interminable.¹¹ (Perhaps it's unsettling that nothing can erase the blame deserved for past

¹¹ See, e.g., Clarke [2022] and Fischer [2014].

wrongs, but there's an uplifting corollary: the praiseworthiness of our actions endures, too.)

Blame is interesting because in some sense, it's more fundamental than resentment. While resentment might be epistemically prior to blame because it judges that someone has done something blameworthy, it is less fundamental than wrongdoing and blame in the sense that it *responds* to them. Accordingly, it's not uncommon to hold that there can be blame without resentment. For instance, Coleman and Sarch [2012] propose that after enough time has passed, resentment for past wrongdoing can fade while blame endures.¹² Obviously, what I've said above explains why resentment also endures. But I mention these views because it's *even more* plausible that blame endures eternally.

There are two cases in which one might attempt to resist enduring blame. The first occurs when a wrong occurs, and something is done to repair that wrong or restore the victim. For instance, consider Kyrie, a self-centered basketball player who commits a series of misdeeds: he causes a young team to implode then refuses to play thus forcing a trade, arrives at a new team and again fails to play, posts links to anti-Semitic videos then doubles down and refuses to apologize, and ultimately leads to the destruction of that second team as well. He is blameworthy for all of these actions. However, time passes and he apologizes for his past behavior. Furthermore, he takes steps to make amends—e.g., by learning more about the history of anti-Semitism and donating to the Anti-Defamation League—and his behavior on his new team suggests that he has reinvented himself into a model teammate. In this type of case, it's tempting to release Kyrie and other similarly situated people from judgments of blame.

However, future reparative efforts, even if successful, don't erase blameworthiness for past actions. For instance, suppose that in response to Kyrie's changed ways, someone demands that sports commentators who remain critical of Kyrie's past actions retract their criticisms and apologize. If blameworthiness can be erased, then these sports commentators ought to apologize for recent criticisms of those past actions because there is nothing worthy of criticism. But what Kyrie did was obviously criticizable and still is. What's true is that sports commentators ought not fixate on just the blameworthy aspects of Kyrie's behavior, but also give proper due to all he's done since that is praiseworthy.

In more extreme cases, this urge to think of blame as something that can be erased leads to pragmatically perverse thinking. Consider, for instance, the "Green Dot" programs

¹² Interestingly, they go on to argue that this is why in certain situations, after a long enough time has passed, it may be inappropriate to prosecute and punish people even though they are still blameworthy.

at some college campuses. These programs encourage people to respond to instances of harassment and assault (*i.e.*, “red dots”) by intervening or aiding the victims, thus creating “green dots” that supplant the red dots. While encouraging intervention is admirable, the stated goal of “creating a Green Dot campus” betrays the fundamental problem of perceived erasure. A campus full of green dots is morally inferior to a campus with no dots at all. Even if you think green dots succeed in making victims whole—and even if you think it is the wrongdoer who corrects the situation—it would be even better for the harassment or assault not to have occurred in the first place. Recognizing that blame is permanent ~~puts~~ focuses our moral efforts on what matters most: not wronging people in the first place. Being blameworthy is a heavy burden to bear, and recognizing the moral permanence of our actions highlights their moral weightiness.

What about cases that are extreme in the other direction, *i.e.*, cases of extremely small wrongs? For instance, suppose that when Bex was a child, she wore her brother’s clothes without asking for permission. Once, while doing so, she accidentally ripped his favorite shirt while hopping a fence. Bex apologized, bought him a new shirt, and stopped wearing his clothing. Years later, when they are both adults with children of their own, Bex discovers that her brother still resents her a little bit and continues to blame her for ripping his favorite shirt. In some sense, it might be strange for her brother to hang onto this resentment decades later and continue to blame her, even if the level of resentment and blame are extremely low, as they ought to be given the mildness of the offense. But there’s a difference, albeit a nearly negligible one, between a past in which Bex wrongs her brother in this way and a past in which Bex never commits this wrong. The endurance of a very minimal level of resentment and blame captures the moral difference between these scenarios.

There’s a second way in which one might try to escape enduring blame. People change, sometimes for the better. In some cases, these changes might be so extreme that they become a “different person” in some relevant moral sense: maybe their character has changed so much that they could no longer commit the type of act for which they are blameworthy.¹³ For instance, perhaps Kyrie has changed so much that he is incapable of being self-centered (or otherwise problematic) and always would put his teammates first. Though rare, this type of transformative change can happen, and when it does, perhaps the new person is no longer the same person and is not blameworthy.

¹³ Chan [2020] makes this point in the context of punishment. Lackey [2020] offers a similar point as part of an argument against long-term punishments.

Khoury and Matheson [2018] offer a less extreme version of this proposal. On their view, a later person is the same person as an earlier person who committed a wrong so long as they are psychologically continuous with them (*i.e.*, that any psychological changes that occur between the earlier and later person are incremental). However, the later person may not be psychologically connected in the right way. For instance, given enough time and enough moral improvement, the later person might have the psychological profile of a moral saint whereas the earlier wrongdoer (with whom they are psychologically continuous) might have the psychological profile of a brutal conquistador.¹⁴ Finally, according to Khoury and Matheson, having the relevant psychological profile is necessary for moral responsibility. So, even if the later and earlier person are strictly speaking the “same person” as the brutal conquistador, the later person with the psychological profile of a moral saint is not responsible or blameworthy for the actions of the conquistador.

Reasonable people can disagree about whether this type of transformation required by this second way of escaping blame is possible. But if possible, this type of case highlights an important point about the objects of blame and our reactive attitudes. Sometimes, the object is narrow: a person *for* a specific action. For instance, Kyrie is blameworthy *for* abandoning his previous teams and praiseworthy *for* his current leadership in the 2024 playoffs. And it’s appropriate to resent and love him for those actions. But sometimes, our reactive attitudes are aimed at the person themselves. When persons themselves are the objects, this possibility of becoming someone new matters for evaluating blameworthiness and the fittingness of resentment. A transformed person, who is not relevantly the same as the person who committed the wrongs, would be neither blameworthy nor the object of fitting resentment. This leads to a natural corrective path for those who have committed wrongs and want to avoid blame. They should attempt to transform by habituating their characters such that they are no longer the type of person who could commit those wrongs.

You may stop here. Section 6 is optional.

6 Forgiveness

Finally, the account I’ve sketched rules out some accounts of forgiveness. In particular, one popular understanding of forgiveness—that it requires the releasing of resentment—becomes impossible, at least if that releasing of resentment must be done fittingly. If

¹⁴ The distinction between psychological continuity and connectedness employed by Khoury and Matheson originates from Parfit [1984].

reactive attitudes endure eternally, then *a fortiori*, resentment endures forever. It's never fittingly released. Thus, either forgiveness (of the kind that requires releasing of resentment) is impossible or it requires unfitting attitudes. Put another way, if forgiveness that requires the releasing of resentment, then it is impossible to forgive without adopting unfitting attitudes.

To be sure, a world in which fitting forgiveness is impossible is unpalatable, as is a world without any forgiveness whatsoever. Perhaps this is too unpalatable to render my view acceptable. Fortunately, my account points to an alternative that permits eternally enduring attitudes, including resentment, while also giving us what we want in a theory of forgiveness.

There are two core *desiderata* we should want out of a theory of forgiveness. First, forgiveness should allow for some sort of reconciliation.¹⁵ We want love and a positive relationship to be possible after wrongdoing. Many assume that having this kind of positive relationship is incompatible with resentment, which explains why the releasing of resentment is such a popular conception of forgiveness. Where there's no longer resentment, there's no longer a compelling reason to refrain from loving or engaging in a relationship. Second, forgiveness must be *uncompromising*. In other words, if resentment is to be released, it should be released for the "right" reasons—not merely prudential ones. As Hieronymi [2001] points out, forgiveness must continue to take the wrongdoing morally seriously. The demand to be uncompromising makes it difficult to see how one can release resentment, since recognition of the wrong makes resentment fitting.

Fortuitously, the story I've told offers an alternative to forgiveness (understood as the releasing of resentment) that also satisfies the two core *desiderata*. I'll refer to this alternative as *forgiveness as non-fixation*, though I'm ambivalent about whether the reader believes it ought to count as a genuine theory of forgiveness. On the one hand, as I'll argue, this alternative satisfies the two core *desiderata*. Furthermore, other theories of forgiveness—most notably Bishop Butler's—do not require the releasing of resentment, so I take myself to be in good company when it comes to resentment-compatible theories of forgiveness. But on the other hand, I concede that many philosophers who work on forgiveness may resist classifying non-fixation as a genuine theory of forgiveness. Ultimately, whether forgiveness

¹⁵ *E.g.*, Strabbing [2020] argues that forgiveness is openness to reconciliation. While I agree that forgiveness must include openness to reconciliation, I disagree that the releasing of resentment is necessary for reconciliation. Bishop Butler makes a similar point when he claims that forgiveness doesn't require giving up resentment, just a particular type of inappropriate resentment.

as non-fixation counts as a “genuine” theory of forgiveness is primarily a linguistic choice. I’m happy for my alternative to be a forgiveness-adjacent practice that better satisfies the two *desiderata*.

Essentially, forgiveness as non-fixation affirms that resentment and its underlying reasons endure eternally. But, so do a host of other attitudes and reasons. This means that one ought not fixate on resentment. Instead, resentment ought to exist alongside and proportionally to other fitting attitudes. For instance, it’s fitting for Kyrie’s teammates and fans to resent him for being a self-centered teammate who puts himself first and causes his team to implode. However, when he later sees the error in his ways and becomes a model teammate, it’s fitting for those same people love him for the way he has turned things around. Further, loving Kyrie for his reformation is compatible with resenting him for his self-centeredness. The central thing to see is that resentment shouldn’t preclude the possibility of love, or any other attitude.

This story accounts for other common judgments about resentment. For instance, it seems strange to resent someone for very small wrongs that occurred in the distant past and have no present effects, as would be the case if Bex’s brother continued to resent her for taking his shirt and accidentally ruining it when they were children. On forgiveness as non-fixation, it’s fitting to hold a very small amount of resentment with respect to a small wrong, but it’s not fitting to fixate on it. Presumably, by the time they are adults, Bex has provided all sorts of reasons for her brother to hold other attitudes toward her. Thus, even though resentment is still fitting, *focusing* on resentment for that past wrong—even if it is of a fittingly small size—is inappropriate if it displaces other more important attitudes.

Wonderly [2021] offers another sort of case in which it appears that there is something inappropriate about holding onto resentment for a past wrong. In this case, Wanda cheats on her partner Victor, with whom she shares a committed, monogamous relationship. Wanda expresses regret, apologizes, makes amends, and ceases having the affair. However, Victor continues to make cutting remarks about her betrayal long after she has recommitted herself to the relationship. This seems wrong. In Wonderly’s presentation of the case, Victor claims that he forgives Wanda even as he continues to make these cutting remarks.¹⁶ Obviously, I’m not going to say that the thing that makes his clearly enduring resentment of Wanda weird is that in forgiving her, Victor is supposed to have released his resentment. Even if Victor did not claim to forgive her, his behavior is still inappropriate. What explains the

¹⁶ Wonderly uses the case in the context of defending a theory of un-forgiveness.

unfittingness of Victor's behavior is his obvious fixation on her affair. It's fitting for him to continue resenting Wanda for cheating on him, but not at the exclusion of feeling other fitting attitudes toward her.

Hopefully, these examples illustrate the way forgiveness as non-fixation operates. Now, we can turn to how it satisfies both *desiderata*. First, holding onto resentment takes wrongs seriously. This is important not only for epistemic and moral reasons, but for pragmatic ones as well. Commonly, people advise others to "let go" of resentment and "forgive" because it will make them feel better. Therapists often recommend this to assault or trauma victims. Some people do in fact feel better by following this advice, and for those people, it may very well be that they have all-thing-considered reasons to release resentment even though doing so is unfitting in the sense we have discussed. But for many, releasing resentment creates a sort of cognitive dissonance because it calls for them to adopt an attitude toward an event or person that is incongruent with their moral evaluation. Something similar can be said for the directed negative emotional responses discussed by McSweeney [2023]. Ignoring the fact that these responses are warranted treats trauma reactions as an internal problem with the patient, when in fact the real problem is an unjust world that is external to the agent.

Cases of atrocities make this point even more strongly. Americans are remarkably good at never forgetting, and always resenting, the terrorists who carried out the 9/11 attacks. That resentment, and refusal to forgive, is fitting (though, many subsequent actions—e.g., invading a country that had nothing to do with the attacks and torturing innocent people—demonstrate the danger of fixating, blinding resentment).¹⁷ Yet, many of these same people are extremely eager to release resentment for and forgive past atrocities committed by the United States and demand that others do the same. In addition to betraying both epistemic and moral faults, the unfitting release of resentment creates at least two devastating practical problems. First, the failure to continue to take past atrocities seriously creates apathy toward repairing the wrong and its downstream effects. Second, it gaslights those who do fittingly resent past atrocities and the ongoing wrongs that stem from them. These pragmatic consequences—what stems from the 'communicative' and 'motivational' functions of affective attitudes—demand that we continue to take wrongs seriously by maintaining fitting resentment, and that is exactly what forgiveness as non-fixation provides.

¹⁷ Resentment can be dangerous, which is why even accounts that hold that resentment and forgiveness are compatible (e.g., Bishop Butler's) warn against "excessive" resentment.

Second, not fixating on the resentment and thus remaining open to forming new attitudes allows for the possibility of reconciliation. Resentment for wrongdoing doesn't preclude compassion, love, or any other attitudes that might form the basis of a positive relationship. For instance, suppose Kyrie apologizes to former teammates and genuinely reforms. This reformation doesn't erase the self-centered, destructive behavior or the resentment that fits it. Those wrongs are permanent fixtures in the universe. But what Kyrie does after provides new reasons to embrace him—reasons that don't diminish, and are not diminished by, his previous bad behavior. Non-fixation, or the ability to appreciate new reasons and form new attitudes, provides the seeds of reconciliation and satisfy what we want from a theory of forgiveness.

Some might complain that non-fixation doesn't constitute genuine forgiveness, and that the brand of reconciliation it recommends is instead more akin to "moving on".¹⁸ The sort of reconciliation worth having is the sort built on the releasing of resentment, or so they claim. This reaction stems from a hardcore commitment to forgiveness as requiring the releasing of resentment. If this is the only sense of forgiveness one countenances, then forgiveness is not fitting, and we should not forgive. But when forgiveness is understood in the non-fixation sense, remaining resentful is not only fitting, but also compatible with reconciling.

Indeed, especially in interpersonal cases, there are steps one can take to provide opportunities to create new reasons that might underlie positive attitudes. Take the Biblical example of Jesus asking Peter if he loves him three times, which mirrors the three times Peter denies knowing Jesus during Jesus's time of need. Jesus doesn't have to do this; it would be permissible for anyone in this situation to have nothing further to do with Peter. (And, assuming they did not unfittingly fixate on the three denials, they might qualify as having forgiven Peter on my view.) But by giving Peter the chance to thrice affirm his love, it's clear that Jesus is not only open to reconciliation (which is compatible with continued resentment), but actively seeks it. He gives Peter the chance to create new reasons that can form the basis of a loving, positive relationship.

Thus, at a minimum, forgiveness as non-fixation requires making space for fitting positive attitudes that coexist alongside resentment. That's enough for forgiveness to count as both uncompromising and reconciling. But one can also go beyond this by attempting to create opportunities for positive attitudes to develop. This allows for a deeper sense of

¹⁸ Thanks to Jada Strabbing for raising this objection.

reconciliation, perhaps one that aims at strengthening a relationship. This creation of new opportunities is supererogatory, and thus allows a more maximal sense of forgiveness that is also elective.¹⁹ For those of us who find ourselves in need of forgiveness, non-fixation fulfills what we need while the supererogatory creation of new reasons gives us hope for what we want.²⁰

Rebecca Chan
San José State University
rebecca.s.chan@sjsu.edu

References

- Allais, Lucy [2013]. "Elective forgiveness". *International Journal of Philosophical Studies* **21**: 637–53.
- Arendt, Hannah [1958]. *The Human Condition*. Chicago: University of Chicago Press.
- Arpaly, Nomy [2005]. "How it's not "just like" diabetes: Mental disorders and the moral psychologist". *Philosophical Issues* **15**: 282–98.
- Butler, Joseph [1726]. *Fifteen Sermons Preached at the Rolls Chapel: A Dissertation Upon the Nature of Virtue*. London: G. Bell & Sons.
- Chan, Rebecca [2020]. "Punishing the wrong person". *Public Affairs Quarterly* **43**: 124–41.
- Cherry, Myisha [2020]. *The Case for Rage*. Oxford: Oxford University Press.
- Coleman, Jules and Sarch, Alex [2012]. "Blameworthiness and time". *Legal Theory* **18**: 101–37.

¹⁹ Allais [2013] compellingly that forgiveness is 'elective,' or such that it's permissive to forgive or not. While my account of non-fixation is not elective, the more robust option of creating opportunities for new reasons is, and introduces an element of electiveness into this account of forgiveness.

²⁰ This paper is dedicated to Anand Vaidya, a friend whose qualities will never fail to fittingly inspire those who knew him. I'm eternally grateful to all the people who gave helpful feedback on this paper, including Michaela McSweeney, Donatella Donati, an anonymous referee, the participants at the *LanCog Summer Metaphysics Workshop*, the *Responsibility and Reactive Attitudes Over Time Workshop*, the *Rocky Mountain Ethics Congress*, and the *East-West Philosophers' Conference* (especially Derek Lam, Hannah Kim, and Yarran Hominh). And, I appreciate the encouragement from and conversations with Noah Lammers, Ned Markosian, David Friedell, Tony Aumann, Andrew Khoury, and Mike Rea; I doubt I would have undertaken this project without it.

- D'Arms, Justin and Jacobson, Daniel [2000]. "The moralistic fallacy: On the 'appropriateness' of emotions". *Philosophy and Phenomenological Research* **61**: 65–90.
- Fischer, John M. [2014]. "Comment on blog post. Blameworthiness over time. Flickers of freedom". https://philosophyonline.typepad.com/flickers_of_freedom/2014/09/blameworthiness-over-time/comments/page/1/#comments.
- Hieronymi, Pamela [2001]. "Articulating an uncompromising forgiveness". *Philosophy and Phenomenological Research* **62**: 529–55.
- Howard, Christopher [2022] "Forever fitting feelings". *Philosophy and Phenomenological Research* **107**: 80–98.
- Lackey, Jennifer [2020]. "Punishment and transformation". In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, E. Lambert & J. Schwenkler (eds). Oxford: Oxford University Press.
- Lebens, Samuel [2023]. "Time for some robust atonement". *Faith and Philosophy: Journal of the Society of Christian Philosophers* **40**: 1–23.
- Marušić, Berislav [2022]. *On the Temporality of Emotions*. Oxford: Oxford University Press.
- McSweeney, Michaela [2023] "Maladjustment". *Philosophical Studies* **180**: 843–69.
- Na'aman, Oded [2021]. "The rationality of emotional change: Toward a process view". *Noûs* **55**: 245–69.
- Oddie, Graham [2005]. *Value, Reality, and Desire*. Oxford: Oxford University Press.
- Parfit, Derek [1984]. *Reasons and Persons*. Oxford: Oxford University Press.
- Singer, Peter [1972]. "Famine, affluence, and morality", *Philosophy and Public Affairs* **1**: 229–43.
- Srinivasan, Amia [2018]. "The aptness of anger", *The Journal of Political Philosophy* **26**: 123–44.
- Strabbing, Jada T. [2020]. "Forgiveness and reconciliation", *Australasian Journal of Philosophy* **98**: 531–45.
- Strawson, Peter F. [1962]. "Freedom and resentment", *Proceedings of the British Academy* **48**: 1–25.
- Wonderly, Monique [2021]. "Forgiving, committing, and un-forgiving". *Philosophy and Phenomenological Research* **104**: 474–88.